# Gaurav Tuli – Applied ML Portfolio (Projects 09–16, Concise 2025)

## Preface: Dr. Gaurav Tuli Background

Dr. Gaurav Tuli is a Data & AI leader with 10+ years of experience turning innovative ideas into deployed solutions. He trained in applied machine learning and network science and has led teams to deliver impactful AI projects in healthcare, public health, security, and humanitarian domains. Dr. Tuli has built AI tools adopted by leading hospitals and public health agencies. He co-led a 114-country COVID-19 surveillance collaboration with academic and industry partners, advancing global pandemic monitoring. He won a national innovation award (DHS Hidden Signal Challenge) for developing a novel bio-threat detection platform. He has published over 30 papers (e.g., in PNAS and Lancet Digital Health), and his work has been featured by media outlets like Wired and the BBC. This portfolio showcases a selection of Dr. Tuli’s applied ML projects, highlighting his ability to integrate advanced analytics with real-world problem solving across domains.

## Project 09: Gender Disparity in COVID-19 Clinical Trial Leadership

**1. At a Glance**  
- Analyzed 7,500+ US COVID-19 clinical trials to quantify gender gaps in study leadership.  
- Found women made up ~37% of trial principal investigators, rising to ~49% by 2022.  
- Highlighted underrepresentation of women vs ~50% in other diseases, prompting equity awareness efforts.

**2. The Problem**  
- Urgent pandemic trials initially favored established male-led networks, limiting women’s opportunities.  
- Some research fields had historically fewer women leaders, likely exacerbating COVID-19 leadership gaps.  
- Lack of data left the gender imbalance anecdotal, hindering recognition and action to address it.

**3. The Solution**  
- Compiled a comprehensive dataset of ~8,000 COVID-19 trials (2020–2022) and comparable non-COVID trials.  
- Applied an ML-based tool to infer each investigator’s gender from their name with high confidence.  
- Analyzed gender proportions over time and versus other diseases, using statistical tests to validate differences.

**4. Architecture Overview**  
- **Automated Data Pipeline:** Gathered trial records from ClinicalTrials.gov for COVID-19 and baseline disease areas.  
- **Gender Inference Module:** Used a name-based ML service to probabilistically assign investigator gender.  
- **Data Filtering:** Excluded trials lacking listed investigators or without confident gender classification.  
- **Analytic Engine:** Computed female-leadership percentages by time period and ran cross-disease comparisons.  
- **Advanced Analysis:** Linked leader gender to trial participant gender and segmented outcomes by trial type.

**5. Results and Impacts**  
- Women constituted ~37% of COVID-19 trial leaders overall, climbing to ~49% by 2022.  
- Female-led trials tended to enroll more female participants, suggesting leadership influences inclusivity.  
- The study spotlighted gender gaps and urged institutions to support women in research leadership roles.

**6. Skills and Tools Used**

| **Data Pipeline** | Automated trial data extraction (Python, APIs) |
| --- | --- |
| **ML Tools** | Name-based gender classifier service (Genderize API) |
| **Statistical Analysis** | Chi-square tests and regression modeling (R, Python) |

**7. Cross-Project Capabilities**  
- **Analytical Rigor:** Demonstrated a robust data-analysis approach applicable to diverse research questions.  
- **Data Insight:** Combined domain expertise with data science, a skill applied across projects.  
- **Equity Analysis:** Experience quantifying representation gaps informs efforts to address bias in other domains.

**8. Published Papers/Tools**  
- Findings published in *Lancet Digital Health* (2023), raising awareness of gender disparities in clinical trials.  
- Reproducible analysis pipeline documented for sharing, enabling further research on trial diversity.

## Project 10: COVID-19 Vaccine Effectiveness vs Emerging Variants

**1. At a Glance**  
- Analyzed global survey data to assess COVID-19 vaccine protection as new variants (Delta, Omicron) emerged.  
- Compared infection rates among vaccinated vs unvaccinated groups during major variant waves.  
- Provided real-world evidence of vaccine efficacy changes against variants, informing booster strategies.

**2. The Problem**  
- Emerging variants threatened to erode vaccine protection, creating need for real-time effectiveness data.  
- Traditional clinical studies lagged in measuring vaccine performance across diverse regions and variant strains.  
- Policymakers lacked timely insight on whether vaccines remained protective, risking delayed booster responses.

**3. The Solution**  
- Leveraged a global Facebook user survey to gather self-reported vaccination status and COVID outcomes daily across 100+ countries.  
- Analyzed millions of responses to compare COVID-like illness rates in vaccinated vs unvaccinated populations over variant waves.  
- Used statistical models to estimate infection risk reduction by vaccine status for each variant, adjusting for demographics.

**4. Architecture Overview**  
- **Data Integration:** Aggregated daily survey responses and aligned them with variant prevalence timelines from genomic surveillance.  
- **Variant Segmentation:** Tagged survey data by time and region to correlate infection rates with dominant variants (e.g. Delta, Omicron).  
- **Effectiveness Modeling:** Employed statistical and ML models to compute odds ratios of infection for vaccinated vs unvaccinated groups per variant.  
- **Dashboard & Visualization:** Built a dashboard to visualize vaccine performance trends over time and across regions as variants arose.  
- **Quality Control:** Weighted survey data for representation and validated model outputs against external case and hospitalization data.

**5. Results and Impacts**  
- Demonstrated high vaccine effectiveness (~85%) against early COVID strains, with moderate declines against Delta and further reduction against Omicron.  
- Real-time analysis flagged waning vaccine protection with new variants, supporting timely public health decisions on booster recommendations.  
- Showed that participatory survey data can rapidly evaluate vaccine performance globally, influencing policy in areas lacking formal studies.

**6. Skills and Tools Used**

| **Data Analysis** | Large-scale survey data processing (Python, R) |
| --- | --- |
| **Epidemiological Modeling** | Vaccine effectiveness calculations (odds ratios, risk reduction) |
| **Data Integration** | Merged survey outcomes with variant prevalence data |
| **Cloud & Collaboration** | Scalable cloud pipeline; coordinated analysis across institutions |

**7. Cross-Project Capabilities**  
- **Rapid Public Health Analysis:** Ability to quickly analyze global data for urgent health decisions (e.g., pandemic response).  
- **Adaptive Modeling:** Experience adjusting models for evolving conditions (new variants) transferable to other dynamic scenarios.  
- **Global Survey Expertise:** Strengthened skills in using large participatory datasets to complement traditional surveillance in various domains.

**8. Published Papers/Tools**  
- Study results published in *Nature Communications Medicine* (2024), detailing vaccine efficacy trends across variants.  
- Findings shared with global health authorities, providing data-driven guidance for booster policies and pandemic planning.

## Project 11: DNS Log Analytics for Malware C&C Detection

**1. At a Glance**  
- Developed a streaming analytics pipeline for real-time detection of covert malware communications hidden in DNS traffic.  
- Identified both fast and stealthy “low-and-slow” malicious DNS patterns that evaded traditional batch detection.  
- Enabled immediate security alerts from DNS logs, greatly improving response time to early-stage breaches.

**2. The Problem**  
- Attackers can hide command-and-control signals in ordinary DNS queries, bypassing standard network defenses.  
- Organizations face huge volumes of DNS data, and manual monitoring can’t catch subtle malicious patterns.  
- Stealthy DNS “tunnels” blend into normal traffic, often escaping detection until significant damage is done.

**3. The Solution**  
- Built a microservice pipeline to continuously ingest and enrich DNS logs (e.g., adding GeoIP and WHOIS context) for live analysis.  
- Implemented multi-scale anomaly detection: short-window analytics to flag immediate threats and long-window analytics to uncover slow, stealthy DNS tunnels.  
- Integrated threat intelligence feeds to flag known malicious domains and deployed a real-time alerting dashboard for security teams.

**4. Architecture Overview**  
- **Modular Microservices:** Independent services handled data ingestion, threat intel updates, anomaly detection, analytics, alerting, and the user interface.  
- **Streaming Pipeline:** A Kafka message broker streamed DNS events between services, enabling scalable, real-time processing of log data.  
- **Tiered Storage:** Combined in-memory caching for instant analysis with a persistent datastore for historical pattern detection over longer windows.  
- **Tiered Deployment:** Services were grouped by function (data ingestion, analysis, storage, presentation) to optimize performance and scalability.  
- **Resilient Design:** The microservice architecture allowed isolated scaling and updates, ensuring robust, fault-tolerant operation of the system.

**5. Results and Impacts**  
- Achieved near-instant alerts for emerging DNS-based threats, vastly reducing attacker “dwell time” in networks.  
- Improved detection accuracy by combining real-time anomaly flags with historical context and threat intelligence (catching both abrupt spikes and subtle slow attacks).  
- Delivered actionable dashboards and alerts that empower security teams to swiftly contain incidents and share new threat indicators with the broader community.

**6. Skills and Tools Used**

| **Streaming & Processing** | Apache Kafka for high-volume, real-time DNS log ingestion |
| --- | --- |
| **Microservices & DevOps** | Containerized services (Docker) for scalable, modular deployment |
| **Data Enrichment** | GeoIP and WHOIS lookups to add context to DNS query data |
| **Threat Intelligence** | Integrated domain blacklists/whitelists via APIs and databases |
| **Visualization & Alerting** | Interactive security dashboard and automated notification system |

**7. Cross-Project Capabilities**  
- **Real-Time Analytics:** Streaming anomaly detection methods here can be reused for live data monitoring in other domains (e.g., social media trends).  
- **Data Fusion:** Combined internal logs with external intel feeds, mirroring multi-source integration techniques used in other projects.  
- **Dashboard Design:** Created intuitive security dashboards – a skill transferable to other analytics platforms (such as public health dashboards).

**8. Published Papers/Tools**  
- Internal Design Blueprint (2025): Architecture and findings documented in an internal presentation for the cybersecurity team.  
- Proprietary Solution: This project was delivered as an internal security tool (no public release due to sensitivity).

## Project 12: Foodborne Illness Surveillance via Twitter Dashboard

**1. At a Glance**  
- Built a Twitter-driven monitoring dashboard to catch local food poisoning outbreaks in real time.  
- Tripled the number of citizen-reported food poisoning cases by engaging ill individuals on social media.  
- Enabled faster restaurant inspections by flagging outbreaks from tweets, significantly improving public health response time.

**2. The Problem**  
- Vast underreporting: Only ~3% of food poisoning cases are ever reported to health departments, leaving most outbreaks undetected.  
- Health agencies rely on consumer complaints, but very few people formally report their foodborne illness to authorities.  
- Untapped signals: Many individuals share food poisoning stories on platforms like Twitter instead of official channels, meaning critical clues were being missed.

**3. The Solution**  
- Developed an automated system to continuously scan Twitter for posts indicating possible food poisoning (using keywords and ML filters).  
- Engaged directly with users who tweeted about getting sick (e.g., via prompts or replies), encouraging them to submit official illness reports to their local health department.  
- Integrated tweet-based alerts into health department workflows by providing a dashboard that mapped suspected cases and facilitated targeted restaurant inspections.

**4. Architecture Overview**  
- **Twitter Data Pipeline:** Set up continuous ingestion of tweets mentioning food poisoning symptoms or related keywords, filtered by location.  
- **NLP Classification:** Employed text processing and machine-learning classifiers to distinguish likely foodborne illness reports from unrelated chatter.  
- **User Engagement Module:** Automated response system that contacts authors of flagged tweets with guidance on how to report their illness to authorities.  
- **Health Dept Dashboard:** A web dashboard for officials that displays a map of tweet-indicated incidents, links to each case’s details, and tracks which have been escalated to inspections.  
- **Multi-City Template:** Designed the system to be easily deployable to different city or state health departments with minimal customization.

**5. Results and Impacts**  
- Deployed across 18 public health agencies (U.S. and U.K.), tripling the volume of citizen-reported food poisoning cases compared to prior reporting rates.  
- Cut detection and response time by ~60%, enabling inspectors to identify and address outbreak clusters much faster than traditional complaint-driven methods.  
- Achieved high precision in identifying true cases: the tweet classification and geolocation approach proved ~91% accurate in pinpointing verifiable foodborne illness incidents in the relevant jurisdiction.

**6. Skills and Tools Used**

| **Social Media Mining** | Twitter API integration for real-time data streaming |
| --- | --- |
| **NLP Classification** | Machine-learning models to flag food-poisoning-related tweets |
| **Geospatial Analysis** | Geocoding tweet locations to map incidents to jurisdictions |
| **Web Dashboard** | Interactive visualization for health officials (maps, alerts, case management) |

**7. Cross-Project Capabilities**  
- **Digital Epidemiology:** Pioneered using social media posts as disease surveillance signals, a method extended in later health monitoring projects.  
- **Public Engagement:** Techniques to prompt and capture user reports via digital platforms were reused in other participatory data projects (surveys, forums).  
- **Real-Time Pipeline:** Expertise in building live data pipelines and anomaly detection translated to subsequent projects in both health and security domains.

**8. Published Papers/Tools**  
- Research Outputs: Three peer-reviewed publications (2018–2020) documenting the methods and impact of Twitter-based foodborne illness surveillance.  
- Operational Tool: The Twitter surveillance dashboard was implemented by multiple city and state health departments, becoming a standard tool for live food safety monitoring.

## Project 13: Identifying Foodborne Illness Reporting Gaps with Yelp Reviews

**1. At a Glance**  
- Analyzed 1.5 million Yelp restaurant reviews for food poisoning mentions to uncover public health underreporting disparities.  
- Identified communities where people report illness online (via reviews) but not through official channels, highlighting equity issues in reporting.  
- Provided data-driven insights to help health officials target outreach and improve food safety reporting in underserved areas.

**2. The Problem**  
- **Reviews vs Reports Gap:** Many diners post about getting sick in online reviews but never notify health authorities, creating an unseen pool of incidents.  
- Potential disparities: Certain demographics or locales may rely on review sites instead of official channels, leaving public agencies unaware of issues in those populations.  
- Missed signals: Traditional surveillance fails to capture these crowdsourced complaints, risking that outbreaks in some communities go unnoticed.

**3. The Solution**  
- Scraped and analyzed millions of Yelp reviews for keywords and narratives indicating foodborne illness experiences at restaurants.  
- Compared the volume and pattern of Yelp-mentioned illness incidents with official food poisoning complaint data to quantify reporting gaps.  
- Highlighted geographic and demographic patterns where many people complain online but far fewer file official reports, enabling health departments to identify where outreach or education is needed.

**4. Architecture Overview**  
- **Data Collection:** Implemented web scrapers to gather Yelp review data across numerous restaurants and regions, building a large dataset of customer feedback.  
- **NLP Extraction:** Used text processing to detect mentions of food poisoning symptoms or complaints within reviews, filtering out unrelated content.  
- **Data Fusion:** Matched the timeline and location of Yelp-indicated incidents with corresponding official public health reports to directly compare unofficial vs. official reporting rates.  
- **Statistical Analysis:** Modeled factors (e.g., neighborhood income levels, restaurant types) associated with high Yelp complaint activity but low official reporting, to understand underlying causes of reporting disparities.  
- **Visualization:** Created charts and heatmaps to illustrate areas of underreporting, helping stakeholders easily grasp the extent and location of the gaps.

**5. Results and Impacts**  
- Revealed hotspots where Yelp reviews indicated numerous illness incidents yet official reports were scarce, signaling critical underreporting in those areas.  
- Demonstrated that lower-income and minority communities had disproportionately more unofficial (online) complaints relative to official reports, highlighting equity and trust gaps in public health reporting.  
- Armed health authorities with concrete insights to focus education and outreach in communities with high “silent” complaint activity, with the goal of improving reporting rates and food safety interventions.

**6. Skills and Tools Used**

| **Web Scraping** | Automated extraction of Yelp review data (Python, BeautifulSoup) |
| --- | --- |
| **NLP & Text Analysis** | Keyword filtering and content analysis of review text |
| **Data Integration** | Linking crowdsourced review data with official public health databases |
| **Statistical Insight** | Analyzed socio-demographic patterns in reporting behavior |

**7. Cross-Project Capabilities**  
- **Non-Traditional Data Utilization:** Showcased the ability to convert consumer-generated content (reviews) into actionable health insights, a methodology applied in multiple projects.  
- **Scalable Analysis:** Gained experience handling millions of data points and extracting meaningful patterns, skills applicable to any large-scale data challenge.  
- **Equity Perspective:** Experience analyzing data for disparities and gaps informs efforts in other projects to ensure fair and inclusive data-driven interventions.

**8. Published Papers/Tools**  
- Peer-Reviewed Study: Published findings in a public health journal, detailing how Yelp review analysis can expose hidden food safety issues.  
- Conference Presentation: Presented this work via a poster at a national public health informatics conference (2020), sharing the approach with a broader audience.

## Project 14: Privacy-Preserving Global Mobility Analysis for Pandemic Response

**1. At a Glance**  
- Analyzed billions of anonymized human movement records worldwide to understand mobility patterns during public health crises.  
- Applied differential privacy techniques to share aggregate mobility insights without exposing individual data.  
- Supported epidemiological modeling and policy-making by providing one of the first global human mobility datasets for research use.

**2. The Problem**  
- Lack of global movement data made it hard to model disease spread and the impact of interventions (travel bans, lockdowns) across countries.  
- Traditional mobility data (travel surveys, limited telecom data) were too sparse or delayed for guiding real-time pandemic response.  
- Policymakers needed timely, privacy-safe insights on how populations move during crises (like COVID-19) to inform decisions, but such data was not readily available.

**3. The Solution**  
- Partnered with tech platforms to aggregate smartphone location data into population mobility metrics, while ensuring individual identities were removed or anonymized.  
- Implemented differential privacy and anonymity checks so that mobility trends could be analyzed and shared without compromising personal privacy.  
- Generated standardized mobility indicators (e.g., distance traveled, visitation frequency, travel between regions) that could feed into epidemic models and help evaluate the effect of policies like lockdowns or reopenings.

**4. Architecture Overview**  
- **Data Aggregation:** Collected massive location datasets from mobile apps and services, consolidating billions of GPS pings into a unified analysis pipeline.  
- **Privacy-Preserving Analytics:** Incorporated differential privacy algorithms (adding calibrated noise to outputs) to ensure no individual’s movement could be re-identified from the aggregated results.  
- **Geospatial Processing:** Computed movement flows and distance distributions by region, using spatial indexing and clustering to characterize typical travel radii and patterns.  
- **Modeling Module:** Fitted statistical models (e.g., power-law decay curves for travel distance vs frequency) to quantify mobility behaviors and how they changed under interventions.  
- **Data Delivery:** Released the processed mobility metrics via public data repositories and interactive dashboards, allowing researchers and policymakers worldwide to access and use the insights in near real time.

**5. Results and Impacts**  
- Released an anonymized global mobility dataset covering dozens of countries, which became a valuable resource for epidemiologists and urban planners (e.g., used in COVID-19 transmission models and policy simulations).  
- Showed how human movement dramatically dropped during early COVID-19 lockdowns and gradually rebounded; these findings validated the impact of interventions and helped authorities plan reopening strategies.  
- Demonstrated a successful collaboration between industry and academia to share big data responsibly for public good, setting a precedent for future data-sharing initiatives in public health and emergency management.

**6. Skills and Tools Used**

| **Big Data Processing** | Cloud-based pipeline managing billions of location records |
| --- | --- |
| **Geospatial Analysis** | Spatial algorithms for mapping and quantifying movement patterns |
| **Privacy Engineering** | Applied differential privacy techniques to protect individual data |
| **Statistical Computing** | R and Python for modeling mobility trends and validating results |
| **Collaboration** | Coordinated with tech companies and research teams for data sharing |

**7. Cross-Project Capabilities**  
- **Large-Scale Data Mastery:** Handling multi-billion record datasets sharpened skills useful for any big-data project (from national security to large health systems).  
- **Privacy by Design:** Experience implementing privacy safeguards (like differential privacy) translates to managing sensitive data in healthcare and other domains.  
- **Cross-Sector Collaboration:** Coordinating with corporate data providers and academics in this project mirrors partnerships leveraged in other initiatives (big tech, government, and research).

**8. Published Papers/Tools**  
- Global mobility findings published in *Nature Human Behaviour* (2020), providing novel insights into worldwide movement patterns (co-authored by Dr. Tuli).  
- An anonymized mobility dataset was publicly released with the study, enabling other researchers and policymakers to utilize global movement data for pandemic planning.

## Project 15: Real-Time Gun Violence Surveillance Platform

**1. At a Glance**  
- Built a real-time platform aggregating Twitter, news, and search data to map gun violence incidents across the United States.  
- Provided up-to-the-minute, state-by-state data on shootings to inform the public and support evidence-based policymaking.  
- Innovatively used digital “exhaust” (social media and web data) to fill gaps in official gun violence reporting and awareness.

**2. The Problem**  
- Official gun violence data was delayed and fragmented, hindering timely awareness of ongoing incidents.  
- The public lacked an accessible, up-to-date source for localized gun violence information and trends.  
- Policymakers had insufficient interim data between official reports to gauge the effects of policy changes on gun violence.

**3. The Solution**  
- Developed an interactive web dashboard that combines multi-source data (social media posts, news reports, and Google search trends) on gun incidents in near real time.  
- Implemented machine learning to categorize gun-related tweets (e.g., distinguishing incident reports vs. discussions) and to classify news articles by type of incident.  
- Geocoded all data to the state level and integrated official metrics (e.g., gun death rates, background check stats) to provide contextual insights alongside the live incident feeds.

**4. Architecture Overview**  
- **Multi-Source Ingestion:** Pipelines continuously pulled data from Twitter (streaming API for gun-related keywords), Google News feeds (for relevant articles), and Google Trends (for spikes in gun-related searches).  
- **ML Classification Pipeline:** Automated tagging of content – grouping tweets into four topical categories (e.g., incident, policy, opinion, other) and news articles into seven categories of incidents – for structured analysis.  
- **Human-in-the-Loop Curation:** A small analyst team reviewed the machine outputs to verify classifications, remove duplicate reports, and annotate developing stories, ensuring data quality and reliability.  
- **Geolocation Engine:** Each tweet and news item was mapped to a U.S. state (using text location cues and a geocoding API) so that the dashboard could display state-by-state incident summaries.  
- **Dashboard Front-End:** A public-facing map and timeline interface updated live, showing each state’s recent incidents, trends over time, and comparisons to official statistics, all in one place.

**5. Results and Impacts**  
- Tracked ~540,000 gun-related tweets and ~1,600 news articles per month during 2015–2016, demonstrating the system’s ability to handle high-volume, real-time data.  
- Deployed the public site **Gunviolencemap.org**, giving citizens and officials a live, interactive view of gun violence trends by state (a first-of-its-kind public tool).  
- Project garnered national attention: it was presented at SXSW 2016 and featured by media outlets (e.g., Wired, KQED) for its novel approach, thereby raising awareness and informing public debate on gun violence.

**6. Skills and Tools Used**

| **Data Sources** | Twitter API, Google News RSS feeds, Google Trends data |
| --- | --- |
| **ML & NLP** | Text classification models for tweets and news content |
| **Geospatial Tech** | Google Maps API for geocoding and state-level mapping |
| **Web Development** | Interactive dashboard (HTML/JavaScript) for real-time visualization |
| **Hybrid Curation** | Combined automated classifiers with human review for accuracy |

**7. Cross-Project Capabilities**  
- **Real-Time Social Data Mining:** Techniques for live social data tracking developed here were later applied to health domains (digital disease surveillance).  
- **Reusable Pipelines:** The Twitter ingestion and classification pipeline was repurposed in other projects (e.g., patient experience sentiment analysis in healthcare).  
- **Integrated Data Dashboards:** Experience merging diverse data streams (social media, open data, official stats) into unified platforms is leveraged in projects requiring multi-source analytics.

**8. Published Papers/Tools**  
- Academic Publication: Demo paper published in ACM Hypertext 2017 showcasing the gun violence data platform and its design.  
- Public Outreach: Invited talk at SXSW Interactive 2016, “Using Social Media to Predict Gun Violence,” which disseminated the project’s approach to a broad audience.  
- Live Platform: **Gunviolencemap.org** launched as a public interactive tool (featured by Wired and others), providing an ongoing resource for stakeholders and the community.

## Project 16: Geospatial Dashboard for Humanitarian Demining

**1. At a Glance**  
- Created a geospatial analytics platform that fuses data from multiple landmine detection sensors into a unified risk heatmap.  
- The interactive dashboard highlights likely landmine hotspots, enabling demining teams to prioritize and track clearance operations more effectively.  
- Designed a modular system architecture allowing new algorithms or sensors to be added without disrupting the whole platform.

**2. The Problem**  
- Fragmented data: Each landmine detection sensor (e.g., metal detectors, ground-penetrating radar, drones) produced separate data, making it hard to see the overall mine risk landscape.  
- High false positives from sensors wasted teams’ time on non-threats, underscoring the need for smarter data analysis to filter out “noise.”  
- No real-time geospatial tool existed to help mission planners decide which suspected mine areas to clear first, leading to less efficient use of limited resources.

**3. The Solution**  
- Architected a five-layer microservice system to ingest various sensor data streams, analyze and fuse them, and present the results in a cohesive way.  
- Continuously generated an updated probability heatmap highlighting likely mine locations (“hot zones”), serving as the core decision-support visual for teams.  
- Incorporated field feedback (reports of found or cleared mines) back into the model, so the system learns over time and improves its predictive accuracy.

**4. Architecture Overview**  
- **Layered Microservices:** Separate services handled data ingestion from each sensor type, data fusion and analysis, risk modeling, user feedback integration, and the user interface.  
- **Heatmap Generator:** The analysis engine produced a geospatial probability heatmap of landmine presence, updating in near real time as new sensor readings and field reports came in.  
- **Interactive Map UI:** A front-end map dashboard allowed users to explore high-risk areas, see details on each zone (e.g., sensor readings, last sweep date), and view progress over time as areas are cleared.  
- **Feedback Loop:** A dedicated module took input from demining teams (confirmed mine finds or cleared areas) and fed that data back into the risk models, continuously refining the system’s predictions.  
- **Planning Module:** An operations planning service combined the latest risk map with external data (terrain, accessibility, team locations) to suggest which suspected areas to address next for optimal safety and efficiency.

**5. Results and Impacts**  
- Delivered a complete system design for a next-generation demining decision-support tool, demonstrating how modern data integration can improve mine clearance strategy.  
- Stakeholders in humanitarian demining reviewed the prototype positively, and plans were initiated to pilot the dashboard in a real demining field operation.  
- Highlighted how advanced analytics and AI can significantly enhance safety and efficiency in post-conflict demining, influencing organizations to invest in smarter demining technologies.

**6. Skills and Tools Used**

| **Sensor Data Fusion** | Combined outputs from multiple detection sensors (metal detectors, GPR, drones) |
| --- | --- |
| **Microservice Architecture** | Modular system design for ingestion, analysis, feedback, and visualization |
| **Geospatial Analytics** | Heatmap generation and spatial risk modeling (GIS tools) |
| **Continuous Learning** | Integrated field feedback loops to update risk models in real time |
| **UI/UX Design** | Interactive map dashboard for mission planning and progress tracking |

**7. Cross-Project Capabilities**  
- **Modular Architecture:** The five-layer microservice approach exemplifies a scalable design pattern Dr. Tuli applies in various domains (from healthcare dashboards to security analytics).  
- **Real-Time Geospatial Analysis:** Skills in creating live risk maps here transfer to other projects involving spatial data (e.g., epidemic spread mapping, environmental monitoring).  
- **Continuous Improvement Loop:** Experience building feedback loops (learning from new field data) is leveraged in other ML systems to ensure they adapt and remain accurate over time.

**8. Published Papers/Tools**  
- Design Documentation: Produced an internal white paper and architecture presentation detailing the demining dashboard design (2025).  
- Early-Stage Project: This was a prototype/plan for humanitarian organizations, so no public release or academic publication was made at the time of writing.